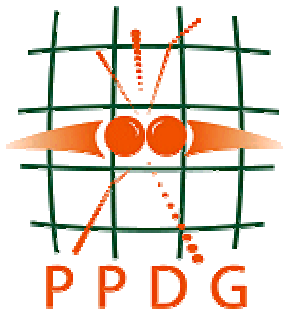
Particle Physics Data Grid Collaboratory Pilot

Quarterly Status Report of the Steering Committee, January - March 2002

2 May 2002



1 Project Overview

1.1 Highlights

PPDG was present at the SciDAC kick off meeting in Washington. We presented a talk and poster session and participated in two panel sessions. The meeting was valuable as a means of learning about other SciDAC projects and discussing ideas and work of mutual interest.

A proposal was submitted to DOE, and accepted by the MICS office, to extend the PPDG project to provide essential work in understanding and interfacing the Site Security infrastructures to the emerging Grid Authorization and Accounting infrastructure.

The PPDG collaboration meeting[1] in March was held co-located with the Global Grid Forum in Toronto. This provided many of us excellent opportunities to learn about and start to contribute to the work of the GGF.

Much progress was made in collaboration with the DOE Science Grid on the issuing and use of DOESG Certificates. D0 demonstrated the first intercontinental use of the DOESG certificates by transferring files between the US and UK using GridFTP integrated into the SAM system. The DOE SG Certificate Authority was subsequently accepted as a "trusted CA" by the European Data Grid, and the US Experiment grid testbeds (CMS, ATLAS, D0).

PPDG teams and collaborators have made significant steps in bringing acceptance of Grid ideas and services to the attention of the experiment mainstream physicist groups. In the experiments: the CMS Grid Integration team agreed upon a comprehensive Grid deployment plan; STAR planned integration of the Grid data transfer and storage management services; D0 added remote analysis and international sites to their Grid deployment and development group; ATLAS plans to integrate MAGDA into the data challenge software of the experiment; BaBar's European sites adopted a strategy of looking at the EDG software suite integrated with the SLAC SRB implementation for dataset replication; JLAB's QCD data replication web portal services will be deployed at a second site (MIT) in the next month or so.

1.2 Project Management and Organization

Torre Wenaus has accepted the position as Applications Area Coordinator for the LHC Computing Grid Project. John Huth, US ATLAS computing and software coordinator, has stepped up to take the ATLAS Team Lead position on PPDG.

1.3 Plans for the next Quarter

A series of internal reviews of the current project activities is planned for April and May. These will provide us an opportunity to survey the results of the project, the perceived benefits from, and costs of the project collaboration.

A two day focus meeting on "Grid Job Scheduling and Management" in May has as a goal to understand the needs and status of the Experiment grids for these services, and plan the work in this area for the next year.

A full day face to face Steering Committee meeting on May 15th is aimed to provide input to the next years of the PPDG vision and goals.

[1] URLs for meetings' agenda and presentations are in Appendix, section 5.2

1.4 Summary of progress on Common Services Development, Integration and TestBeds

1.4.1 New Documents

Document below are posted at http://www.ppdg.net/docs/documents_and_information.htm.

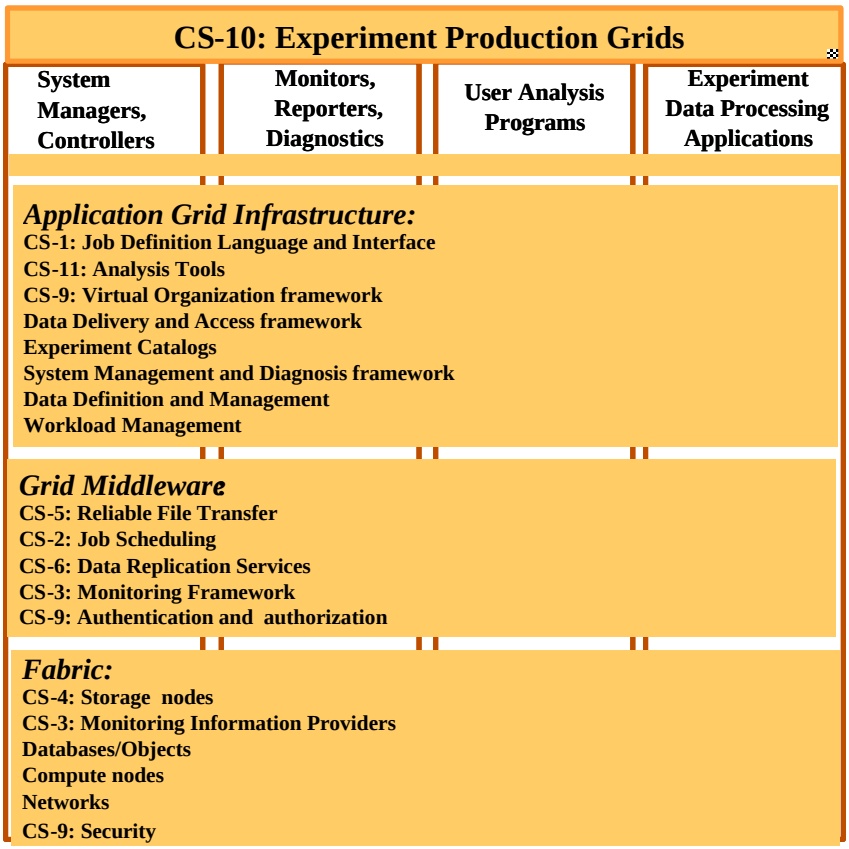
Doc #	Document name, authors	Date
PPDG-14	Oracle Activity for the SAM Database Servers, G. Garzoglio, J. Weigand	3/02
PPDG-13	Cross reference of PPDG CS and EDG WP groups, J. Gieraltowski	2/02
PPDG-12	GridFTP update, Globus.	2/02
PPDG-8	Data Grid Implementations - Comparison of Capabilities, R.Moore et al	updated 2/02

1.4.2 Kick off of CS-11

A new common services cross-cut project was initiated ” Interfacing and Integrating Interactive Data Analysis Tools with the Grid and Identifying Common Components and Services”. Interaction and integration of Physics Analysis software with Grid services are an essential ingredient of an experiments overall data processing requirements. It is expected that work in this area will ramp up and continue for the remainder of PPDG, and will encompass several sub projects such as dataset specification and selection, grid enabled software analysis tools, etc. Several experiments have relevant developments underway.

1.4.3 The PPDG Scope –Experiment Production Grids and Common Services

The adjacent chart categorizes functional components of PPDG experiment production grid systems, show that the End-to-End applications use services spanning architectural layers, and indicate the multiple User and System Applications that make up Experiment Production Grids. Each Experiment Production Grid must deliver worth, act, be manageable and viewable as an integrated whole. Since the scope of PPDG includes deploying user and experiment applications and production grids, the project must concern itself with at least an understanding all required services. Those that require more focused work within the project have, or may have in the future, identified Common Services (CS) activities.



During the first year of this phase of PPDG, we have not attempted to come to agreement on a precise definition of each of the terms. The Experiments and collaborating Computer Science groups are working in several areas simultaneously in order to meet the requirements of the experiments and achieve working systems that can be used to give feedback and better understanding of the precise breakdown of services, common architectural components and interfaces that might usefully be extracted and defined.

1.5 Year 1 Status Update

Project Activity	Experiments	Yr1	Any Changed Status 4/7/02
CS-1 Job Description Language – definition of job processing requirements and policies, file placement & replication in distributed system.			Will be defined in May Focus meeting.
CS-2 Job Scheduling and Management - job processing, data placement, resources discover and optimization over the Grid			Will be subject of May Focus meeting.
CS2-1 Pre-production work on distributed job management and job placement optimization techniques	BaBar, CMS, D0	X	CMS – integration of prototype into simulation production well underway on the US CMS testbed D0 – SAM in use by the collaboration.
CS-3 Monitoring and Status Reporting			
CS3-1 Monitoring and status reporting for initial production deployment	ATLAS	X	Joint monitoring group GIIS server established at BNL. Extension of fabric monitoring deployed.
CS3-2 Monitoring and status reporting – including resource availability, quotas, priorities, cost estimation etc	CMS, D0, JLab	X	CMS – US testbed starting to use MDS and understand how to interface/integrate with monitoring frameworks and displays D0 – extension of analysis of database and log file statistics gathered and published weekly.
CS-4 Storage resource management			
CS4-1 HRM extensions and integration for local storage system.	ATLAS, JLab, STAR	X	STAR – deployment of prototype into experiment production started. JLAB – deployment of prototype integrated with production HSM (Jasmine)
CS-5 Reliable replica management services			
CS5-1 Deploy Globus Replica Catalog services in production	BaBar,	X	Rewrite of prototype using SRB and extended BaBar catalogs, in test.

CS5-2 Distributed file and replica catalogs between a few sites	ATLAS, CMS, STAR, JLab	X	GDMP V3.0 released. ATLAS testing GDMP from Magda. CMS US testbed using GDMP V2.1 and soon migrating to GDMP V3.0
CS-6 File transfer services			
CS6-1 Reliable file transfer	ATLAS , BaBar, CMS, STAR, JLab	X	All using GridFTP in production or in test. D0 has integrated GridFTP into SAM.
CS-7 Collect and document current experiment practices and potential generalizations	All	X	New documents listed above

1.6 Interactions with other Projects and Activities

PPDG is collaborating more closely with its peer US Physics Grid Projects – GriPhyN, iVDGL. GriPhyN and iVDGL will adopt the use of DOE SG certificates and we are providing help and support through the PPDG RA and CP. PPDG is working with the Virtual Data Toolkit team to provide GDMP, Certificate signature files, and other configuration information as needed. PPDG will collaborate with GriPhyN and iVDGL on the support infrastructure for Virtual Organizations and the evaluation of the Globus Community Authorization service.

PPDG is also contributing as part of the High Energy Physics Intergrid Joint Technical Board and US-European Grid Logical Uniform Environment[2] (GLUE) effort to provide and promote interoperability between the European and US physics Grid efforts. Three team Leads, Ian Foster, Lothar Bauerdick, John Huth, of PPDG are members of the LHC Computing Grid Project[3] Software Coordination Steering Committee (SC2) and members of the executive team (Miron Livny, Ruth Pordes) are external members of the Project Execution Board.

Doug Olson attended the all-hands meeting of the Scientific Data Management ISIC[4] meeting[5] in March, representing HENP applications.

The effort to define a common schema for grid information services as part of the GLUE activity is making important progress in helping ensure interoperability of US and European grids.

2 Project Activities

2.1 GDMP (CMS-DataGrid-Globus)

In this quarter, two GDMP releases (GDMP 2.1 and GDMP 3.0, respectively) have been made and several new features have been added. Both releases and corresponding activities are described below.

In the File Replication Meeting at JLAB (10 January 2002) the status of GDMP was presented and discussions on integrating GDMP into other PPDG software systems were done, in particular, with

[2] http://.hicb.org/glue
[3] http://lhcgrid.web.cern.ch/LHCgrid/
[4] http://sdm.lbl.gov/sdmcenter/
[5] http://sdm.lbl.gov/sdmcenter/pub/allhands.htm

MAGDA. A follow up meeting with MAGDA took place at CERN in March 2002 where several of the open issues were discussed and mostly met in GDMP release 3.0.

Although GDMP 2.0 has already been used successfully in the EU DataGrid testbed, for GDMP 2.1 (released in the beginning of February 2002), additional features have been added to meet the requirements to also use GDMP in a Data Grid testbed where people do not have direct access to the Storage Elements were GDMP is installed. In addition, GDMP has been upgraded to be used with Globus 2.0 beta (Globus release) and Globus 2.0 beta 21 (EDG release).

2.1.1 GDMP version 2.1

The major new features of GDMP 2.1 were to have a preliminary multiple VO functionality: in particular, configure_gdmp can be used to configure several VOs on one StorageElement and preliminary multiuser support is gained via file access group permissions. A few existing bugs have been fixed and a new User Guide with detailed examples for the EU DataGrid Testbed has been done.

On 1 March 2002 the first EDG project review took place and GDMP was successfully used in the DataGrid demonstration.

GDMP 2.1 is also part of the GriPhyN Virtual Data Toolkit (VDT).

2.1.2 GDMP version 3.0

In parallel to GDMP 2.1, GDMP 3.0 development was on going and GDMP 3.0 was released in the first week of April 2002.

This release is supposed to be the last major GDMP release of this type with all GDMP code in C++. GDMP 3.0 is supposed to be a stable version that meets several user requirements from PPDG and EDG. The entire list of new features can be found on the GDMP web page at: http://cmsdoc.cern.ch/cms/grid/software.html

In summary, the main new features contain:

- new security model for the GDMP server
- real client-server behavior for GDMP command line tools: The GDMP installation is divided into a full installation (client and server part) and a client installation. This separation is essential in a Data Grid where users do not have access to the Storage Elements where GDMP servers are running.
- GDMP client API in C++
- real multi-user multi-VO support with a single GDMP installation
- new client commands
- very detailed documentation and examples in the GDMP User Guide

2.2 D0 Job Management (D0-Condor)

D0 has started work on architecture and principal design of the “D0 Grid” job management. The principal components are the resource information collector and the request broker. D0 has analyzed the possibility of basing the initial job management architecture on Condor facilities, most notably, on Condor-G and Condor’s Match-Making Service (MMS). The D0 and Condor teams met at the University of Wisconsin to discuss how such an architecture can be accommodated.

We have identified two areas where Condor has to be extended, these are: extension of Condor-G as to enable the determination of the target Gatekeeper by the MMS, and a “callout” capability in MMS to evaluate and externally executed function. Work is underway on the Condor side to implement these functionalities in a prototype. On D0 side, work continues with exploration of the D0 Condor-G testbed on job submission and monitoring.

In the next quarter, we plan to test the above Condor extensions in the D0 testbed, as well as study how Condor can be useful for D0 job monitoring.

2.3 CMS-MOP (CMS-Condor)

The goal for this quarter was to get MOP running current CMS production activities on the US CMS Grid Testbed. Through collaboration between Fermilab and Wisconsin, and with help from the teams at all US CMS prototype Tier-2 sites, UCSD, Caltech, and U. Florida, the goal was met.

The efforts at Fermilab focused on updating MOP for Globus 2.0 and integrating with the current CMS production environment. James Amundson worked on the Globus 2.0 updates, including support for GridFTP. He also added explicit support for GDMP publishing operations. He also added various other convenient improvements, including the ability to remotely install the CMS software packages, DAR files. Greg Graham updated the CMS production software, IMPALA, to work well with MOP. Peter Couvares and Jame Amundson also contributed to this effort. GDMP installation and bug fixing was supplied by Shahzad Muzaffar. The Fermilab team worked closely with the Wisconsin team on the transition from James Amundson's lead to Peter Couvares's lead.

Wisconsin focused on continuing collaboration with Fermilab and the US-CMS Testbed to design, develop, deploy, and test the IMPALA/MOP software and its many underlying components, including Globus, Condor-G, Condor, GDMP, BOSS, and the CMS application software. Over the quarter, this effort consisted of considerable design and development work, and then transitioned into ongoing deployment and debugging on the actual testbed. By the end of the quarter the MOP team successfully demonstrated distributed production of simulated CMS events using the CMKin and CMSim programs on the testbed.

2.4 STAR-DDM (STAR-LBNL/SDM)

The focus of STAR-DDM during this quarter was to attempt to use the HRM 3.0 software that was installed at BNL and NERSC in the previous quarter for STAR's large-scale data replication needs. The first step in this process was to transfer knowledge of how HRM works from the developers (SDM) to the users (STAR-LBNL). This was completed in January along with a transition from the Globus 1.1.3 alpha version to the Globus 2.0 beta release. In February STAR collaborators continued testing the performance of the HRM's for data transfer. The HRM's themselves performed well – sourcing data out of HPSS at BNL via pftp to a cache disk at BNL, transferring it to a cache disk at NERSC via GridFTP, and finally sinking it back into HPSS at NERSC via hsi. Unfortunately a problem with GridFTP (also seen in initial tests by the SDM group) resulted in reliability and bandwidth problems. The problem was that GridFTP would "hang" after some number of successful transfers, and the larger the net transfer rate (through concurrent transfers) the sooner GridFTP would hang. This rendered the performance of HRM/GridFTP too slow and/or too unreliable for STAR data transfer, and this was reported at the PPDG collaboration meeting in late February. After extensive communication between the SDM group and Globus people the problem was reproduced and fixed by the Globus team. The fix will be available as part of the official Globus 2.0 release which is expected in April 2002. In the meantime work has begun on a "replica coordinator" to run the HRM's. In its first version this code uses globusrun to query STAR's mysql file catalog at BNL (behind the firewall) from NERSC. Query results are transferred back to NERSC via GASS, formatted for use by the HRM at NERSC and the transfer can be initiated. Work in the upcoming quarter will focus installing and testing Globus 2.0 when it is released, further development of the replica coordinator to automate data transfer and developing replica catalogs.

2.5 JLAB-Replication (JLAB-SRB)

Jefferson Lab and SRB are continuing to work toward defining a web services interface to a file replication system as part of a strategy to provide a common layer to multiple storage systems. The goal in the first year is to define and implement a common interface to replication services provided by SDSC's SRB software and Jefferson Lab's JASMine software.

An important milestone will be the specification of the web services using Web Services Definition Language (WSDL). Specifying this WSDL document requires agreeing on

- System decomposition into network objects (web services)
- Functional decomposition into specific operations (function and argument names and semantics)

Discussions of these topics were held at the Reliable Replication Workshop hosted at Jefferson Lab, and also at the Global Grid Forum 4 meeting in Toronto. These discussions have continued by email during this three month period, as both JLab and SRB have continued to develop prototype web services for their systems. These prototyping efforts are helping to guide our formulation of the WSDL for replicated data grids.

We have been looking at the issue of development of a WSDL services interface that can run across both the JLab and SRB systems. We demonstrated WSDL interfaces at the Global Grid Forum meeting on Feb 20. Chip demonstrated a WSDL interface for JLab, and Arcot Rajasekar demonstrated a WSDL interface for the SRB. Since then we have discussed the unification of the two WSDL interfaces. The challenges as discussed at the GGF4 meeting are:

- identification of which parameters are managed externally to the WSDL call, which are passed as service calling arguments, and which are determined dynamically within the server software.
- asynchronous versus synchronous access
- collection or community ID used to store data versus user-owned data
- use of global file name versus local Site URL for referencing files
- authentication mechanism

We have worked on a joint description of the interface that would create a representative WSDL service for data replication. We recognize that the publication of the Open Grid Services Architecture will impact our attempt at defining a WSDL interface. We also recognize that an alternate interface may be under development within the Globus community. I believe that the most urgent architecture consideration is the first point that was listed, namely the development of an understanding of how much of the state information will be managed by the grid software for each type of service.

2.6 ATLAS distributed data manager, MAGDA (ATLAS-Globus)

The principal goal of the Magda project for the period was the application of Magda in the ATLAS Data Challenges, supporting it in production, and feeding back experience into ongoing development. A production ready version of Magda was released in early December 2001, conforming to the original ATLAS Data Challenge 0 schedule. The DC schedule has seen modifications and delays since then, however. DC0 production is complete and all results have been successfully cataloged by Magda and made available to the community via Magda access and replication tools. The principal DC0 deliverable of a complete processing chain within one release remains incomplete at this date. Despite this, the Collaboration has moved forward with an aggressive plan to adhere to the DC1 schedule commencing in April to deploy and exercise distributed production on about 18 sites worldwide. Magda was accepted for use as the means of data replication and cataloging for this exercise. Wensheng Deng, since March 1 a full-time PPDG (BNL) developer focusing on Magda, worked closely with the Data Challenge team to support this usage. First production is scheduled for May.

DC1 ‘phase 0’ as the distributed production exercise is called provides us the opportunity to exercise the replication functions (as distinct from the cataloging functions) of Magda in large scale production for the first time. To this end, and also to satisfy the needs of small scale replication users below, the replication mechanisms were made ‘user accessible’ as opposed to ‘experts only’ by improving the information, documentation and task control available to users via the web interface. Step by step procedures accessible to end users were developed, tested and documented. The replication mechanism was also extended to more flexibly support disk to disk transfers (not involving mass stores) between sites.

We also responded to a usage request from the liquid argon calorimeter group at BNL to use Magda for near real time management of bench test data acquired on dedicated DAQ systems and archived on the Tier 1 mass store. This provided us with a new use case we proceeded to implement, involving a dynamically changing input data set (source location content changes continuously as new data files are acquired and

registered) to be replicated. Magda was extended to support replication of such dynamic data sets in addition to static ones, and this usage is now under test.

In light of the growing attention to web services, as a learning exercise on SOAP technology we implemented a SOAP testbed and explored its applicability for an implementation of Magda's "SQL accelerator" by which batch database transactions are transmitted in bulk over the network and executed by a CGI-driven trigger (which would be replaced by SOAP). The technology fits the application well, but given other priorities we are not proceeding with a re-implementation at this time.

In a January PPDG focus meeting on data management we presented the results of a review of GDMP's appropriateness to use by Magda for publish/subscribe replication. We presented a 'GDMP wish list' which has been fully addressed in the forthcoming GDMP Version 3 release. Accordingly we plan to proceed with GDMP integration in Magda using V3, to provide a production-oriented publish/subscribe replication service in Magda.

GDMP integration with Magda is a means by which ATLAS is planning to merge its US and EDG based grid activities. Magda's application in the ATLAS Data Challenges attracted the interest of the ATLAS EDG team, and an agreement was reached to make an EDG person (a computer scientist working for Laura Perini at Milan) available to work on Magda and EDG/GDMP integration. The work is expected to begin in April.

Magda usage in ATLAS continued to grow, with the cataloged data volume exceeding 10TB at the end of the period. Participants in the activity during the period were Wensheng Deng and Torre Wenaus (BNL). Alex Undrus provided database and infrastructure support and system administration.

2.7 BaBar Database Replication (BaBar-SRB)

Researchers from both the SRB team and BaBar participate in weekly collaboration meetings to discuss SRB implementation issues within the BaBar environment. A second prototype has been created that uses the SRB system to manage data access, with metadata stored in the BaBar Oracle database. Recently we have been concentrating on the following two issues.

1. Use of the PPDG PKI certificate within SRB for the BaBar project. The SRB has supported the GSI authentication software since SRB version 1.1.7. (The latest version is SRB 1.1.8). To use PPDG certificate(s), both the BaBar and SRB teams are going to build a GSI-enabled SRB server and client software for BaBar. We will then test logins to the SRB at Stanford using PPDG certificate(s). Meanwhile, the SRB team is working with the SDSC security team so that SDSC's Certificate Authority will be able to accept PPDG certificate(s).

2. Since BaBar is thinking of running multiple MCAT catalogs, there is now a need to federate multiple catalogs and provide a uniform name space across the multiple systems. One approach is to use the Globus replica catalog. A second approach is to use a SRB system as the replica catalog. A third approach is to use the ability of the SRB system to create shadow links that support the registration of remote files into the MCAT catalog.

By using shadow links, it is possible to minimize the number of certificates that must be interchanged between the institutions running the independent catalogs. Each site only needs to honor the certificate under which the remote SRB environment operates. Also, each community can impose its own access control lists, governing the permissible use of the data. The community that owns the original data can set access controls on the SRB ID used by the second community. The second community can set access controls for their users on the shadow link that they register. This approach has been discussed with Andy Hanushevsky.

Within the BaBar experiment we are still slowly building the relational database that joins with the Mcat. We had a number of problems (mainly because we did not initially manage to capture all the use-cases which required modifications to the schema). We will use the new Globus 2.0 libraries to build the GSI-enabled Mcat server. Once that is complete we will try to test to see if we can access the Mcat from in2p3 using the CNRS certificates (and similarly at in2p3 where they are also setting up an Mcat catalog). We are setting up an Mcat at in2p3 (initially just for testing) then we will need to create scripts that are capable of bulk-copying metadata information from SLAC to in2p3 and populating the Mcat at in2p3.

3 Cross-cut Activities and Collaborations

3.1 Certificate/Registration Authority

The DOE Science Grid Certificate Authority (DOESG CA) has been operating since January 2002 and the PPDG project is operating a Registration Authority (RA) in conjunction with this CA. The purpose of the CA is to issue X.509 certificates used for identifying people, computers and services carrying out transactions on the grid. The role of the RA is for the initial identity check that the person requesting a certificate is really who they claim to be. The PPDG RA performs this identity check step for people in the PPDG community. A website with information about the operations of the PPDG RA is at http://www.ppdg.net/RA/ and contains links to the CA. To date, about 60 certificates have been issued for people and machines via the PDPG RA.

The policies and procedures by which the CA and RA operate are described in the Certificate Policy & Certification Practice Statement (CP/CPS). The procedures of the PPDG RA are described in an appendix of the CP/CPS. This appendix served as a template for the other RA's operating in conjunction with this CA. Each RA has a member on the Policy Management Authority (PMA) whose role is to decide on the policies and procedures of the CA, and contribute to writing the CP/CPS. In general, the level of reliability of the certificates issued by the CA depend upon the procedures described in the CP/CPS and on how well the CA and RA operations adhere to those procedures. The procedures used for this CA were chosen to be equivalent to those used for the CA's operating in Europe and being used for the EU-DataGrid project. This equivalence of procedures (and numerous discussions) has led to the formal establishment of trust between the EU-DataGrid project and the PPDG project in that the grid testbeds of both projects will accept certificates issued the DOESG CA and the EDG CA's.

At this point in time, the issues surrounding the establishment and operations of a CA and RA for PPDG are resolved and the operation of the PPDG RA will continue until no longer needed. It is likely that, in the long run, organizations like the host laboratories for the large HENP experiments will operate Registration Authorities that are integrated with other user registration procedures and it will no longer be necessary for a PPDG RA to operate for the purpose of issuing certificates to collaborators on those experiments.

3.2 Joint PPDG, Griphyn iVDGL Monitoring Project

3.2.1 Leadership

Brian Tierney from LBNL volunteered to become the new co-lead (with Jennifer Schopf, ANL) for the group. Newly funded by Mary Anne Scott, Brian has been working in the field of monitoring and Grid computing for many years now, including as a lead for the performance effort in PPDG and the lead architect for the NetLogger work.

3.2.2 Talks

The first goal for this group was to define use cases for requirements gathering. We did this by first defining a template, and then by requesting use cases from the experimentalists involved in this effort. To date we have 19 of these covering a wide range of examples from testing a network for stability to evaluating the progress of an application. Jennifer Schopf presented this work at the Internet2 End-to-End Performance Initiative Measurement Workshop in January 2002, which is available from http://www-unix.mcs.anl.gov/~schopf/Talks/internet2_pgmon_2_2002.ppt. This work was also part of the talk she gave for the LCG kick-off meeting in March 2002, which is available from http://www-unix.mcs.anl.gov/~schopf/Talks/info_services_lcg_3_2002.ppt.

3.2.3 Requirements document

Serious work started on developing a requirements document for grid-level monitoring issues. The current draft document is available at http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/ReqDoc/monitoring_requirments.v1.pdf. Both Ruth Pordes (FNAL) and John McGee (ISI) are currently working on this document, and we hope to have a final version prepared by May.

3.2.4 Meetings and Re-scoping

This group met informally as part of the GriPhyN meeting in January, and were tasked by the VDT people tasked us with defining a set of sensors to be deployed in the various application testbeds as part of VDT. However, at the PPDG meeting in February it was realized that that goal was really one of fabric management, and work in that area was being done by many members of the PPDG monitoring community already, and what was needed was a better plan for how to interface these various fabric-level monitoring systems to a grid-level monitoring system, such as the Globus Monitoring and Discovery Service (MDS). A new version of the charter has been advertised to the group, and is posted online at http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/charter.html.

This re-focusing will involve the use of the currently being defined unified schemas being developed by the Glue-Schema group (discussed below), the development of needed sensors or information providers to allow inter-operable deployment of this information, and a joint GIIS set-up for this group, which has already been taken care of by Dantong Yu, BNL.

3.2.5 Glue Schema work

In early March the JTB started a group to define, publish, and enable the use of common schemas for interoperability between the EU physics grid projects (focusing on EDG and DataTag) and the US physics Grid projects (focusing in on PPDG, GriPhyN and iVDGL). Brian and Jenny are coordinating this effort. The web page for this project is http://www.hicb.org/glue/glue-schema/schema.htm. This work is part of the Grid Laboratory Uniform Environment (GLUE) Phase I task (http://www.hicb.org/glue/GLUE-v0.04.doc).

This effort will provide a basis from which to understand short, medium and longer term needs and definitions, and will encourage coordinated progress, and increased communication between these groups. No one set of schemas are being adopted, rather a new, unified schema is being developed; with the goal of having schemas defined for use in LDAP, SQL and XML.

The first step is to define common schemas to describe Compute Elements (CE), Storage Elements (SE), and Network Elements (NE), to be used by the MDS and R-GMA Grid Information Services. The goal is to have common schemas defined, deployed, and tested in time for the EU DataGrid Testbed 2 release in September 2002. Common schemas for monitoring and notification events are being address by the Global Grid Forum DAMED working group, and will be addressed later by iVDGL and DataTag. Good progress is being made at defining a common schema for the "Compute Element" at this time.

3.2.6 References

Please note monitoring work for individual groups are discussed in other sections of this report. We maintain strong ties to the experiment application testbeds, especially through the efforts of Les Cotrell (BaBar), participation in the ATLAS and CMS testbed meetings, and the D0 work on MDS evaluation. The web page for the joint PPDG/GriPhyN/iVDGL project is located at http://www.mcs.anl.gov/~jms/pg-monitoring/.

3.3 Site Authentication, Authorization and Accounting

The concept of grid computing and the models being developed by PPDG and the numerous other projects and activities are making good progress on middleware and software integration to enable the applications of physics experiments to take advantage of the grid computing model. However, there has not been much involvement of the people administering and providing the computer security of the large computing facilities at DOE laboratories that are being used by the experiments, and will be comprising the major portion of the computing resources in the data grids for these experiments. Since there are significant issues that need to be resolved about how the Grid Security Infrastructure (GSI) is interfaced to the site security already in place at DOE laboratories, PPDG started an effort in this area, called Site Authentication, Authorization and Accounting.

A proposal[6] was submitted to DOE in March for $0.5M funding in FY2002 which will be used assist these efforts at the five laboratories BNL, FNAL, LBNL/NERSC, SLAC, TJNAF. Details of the activities in this cross-cut project are described at http://www.ppdg.net/pa/ppdg-pa/siteaa/ and a mailing list[7] with web archive is established to support communications for this effort. In addition to the laboratories, there are participants in this effort from the Globus project to ensure effective coupling between the middleware developers and the site security administrators. Contacts with the corresponding efforts in Europe in the DataTAG and EU-DataGrid projects has been established and some initial joint discussions were held at a meeting at ANL on April 26, 2002.

3.4 Interactive Data Analysis Tools

Most of the grid activity to date has been focused on large-scale production computing applications which are extremely important for the science missions but do not directly involve a large number of the physicists who are studying the science of the HENP experiments. In order for the effort being applied to data grids to be most effective in enhancing the science and creativity of the scientists involved it is necessary to consider data grids from the point of view of the individual physicist carrying out data analysis for these experiments. PPDG has just begun a cross-cut activity in this area with a charge to "interface and integrate interactive data analysis tools with the grid and to identify common components and services."

A mailing list[8] and web pages (http://www.ppdg.net/pa/ppdg-pa/idat/) has been established. The initial phase of work is to

- identify appropriate individuals to participate in this area, within and from outside of PPDG
- assemble a list of references to white papers, publications, tools and related activities
- produce a white paper style requirements document as an initial view of a coherent approach to this topic
- develop a roadmap for the future of this activity.

The first two of these steps have taken place and the requirements document is now in the process of being outlined. A workshop is being planned to bring people together for concentrated discussions in mid June.

3.5 Collaboration with IEPM, Network Performance Monitoring

Contact: Les Cottrell, SLAC

We have created a web site[9] organized to provide easy access to all aspects of this project. Several new hosts have been added to those monitored by the Internet End-to-end Performance Monitoring – Bandwidth Project (IEPM-BW). We have added pages on documenting procedures for setting up new remote hosts. We added access from the IEPM-BW data to the PingER reporting web site[10].

We have built tools to facilitate and automate the infrastructure management. This includes downloading of code, gathering the remote configurations parameters (OS, cpu speed, code versions, file locations), verifying windows and streams are set correctly. To assist in auditing links that are having problems we put together a reporting tool to identify problems. We review this on a daily basis to identify where we need to focus effort to keep things running. As a result of this, we made several modifications to the measurement tools to improve robustness and recognize more failure modes.

We have defined and added several new tools to the IEPM-BW suite, including: a simple forecasting tool; We have been working to install the Globus infrastructure at SLAC so we can add GridFTP to the IEPM-BW tool suite. We have made some stand-alone measurements, but are still trying to resolve the certificates

[6] http://www.ppdg.net/docs/PPDG-AAA-Proposal.pdf
[7] http://www.ppdg.net/mailman/listinfo/ppdg-siteaa
[8] http://www.ppdg.net/mailman/listinfo/ppdg-idat
[9] http://www-iepm.slac.stanford.edu/bw
[10] http://www-iepm.slac.stanford.edu/cgi-wrap/pingtable.pl?dataset=iperf

issues. We now have certificates for the DoE-SG and Globus and are working on enabling them. We developed tools to measure the throughputs passively using the Cisco NetFlow tools. This enables us to make forecasts for sites to which we are not making passive measurements, and also to compare and validate the active versus the passive measurements[11].

We have installed the latest Web100 instrumented TCP stack on two Linux 2.4 measurement hosts at SLAC. We are starting to use Web100 to understand the throughput performance.

We have made public the IEPM-BW data and provided documentation on how to access it. The data has been requested by CAIDA, ANL, and the Network Weather Service (NWS). BNL are working on providing a more standard way to access the data.

We have provided assistance to 3 major BaBar collaborator sites (RAL, IN2P3 & INFN/Rome) that were experiencing network problems. We documented the cases so multiple people could work on the problems and to share the information as an instructional tool. To assist others, including PPDG folks, in tackling network related problems we put together a web site of case studies related to poor network performance.

We have made presentations on the new measurement infrastructure and results at several meetings including: the DARPA PI meeting in Washington (presentation), the SciDAC meeting in Washington (presentation); the Internet 2 End-to-end Performance Initiative working group at Tempe Arizona (presentation); the Global grid Forum in Toronto (presentation on QBSS); the Internet 2 HENP networking working group in Toronto (presentation on network measurement); the IEPM-PPDG collaboration meeting in Toronto (presentation); visiting DoE program manager at SLAC (presentation on IEPM-BW, presentation on INCITE work); the International Committee on Future Accelerators Standing committee on Inter-regional Connectivity (ICFA/SCIC) at CERN (presentation); at the International Pure and Applied Mathematics (IPAM) workshop on Internet measurements at UCLA (presentation).

We have had several meetings with various collaborators including 2 phone meetings with the PPDG monitoring group; 2 with the Internet 2 End to end Performance Initiative group; a face to face meeting at SLAC with the LBL team; a phone meeting with Rice; face to face meetings with member of the European DataGrid at CERN and Daresbury Lab; face to face meeting with the Rice/INCITE collaborators and the Delaware/Pathload/Pathrate collaborators.

We submitted a paper to the IEEE on the high performance network measurement results and a second on the effects of compression on throughput to the Edinburgh Global grid Forum meeting in July.

Jiri Navratil from the University of Prague joined the IEPM group, He will be working on evaluating/validating tools from other sites including Rice. U Delaware and LBL Doug Chang left; Yolanda Tsang, a PhD student from Rice visited SLAC for a month.

We are in the process of porting the measurement tools to Linux from Solaris. We are also understanding Web100, deciding what parameters are useful to add to the IEPM-BW measurements, and whether and how Web100 can be used to assist application steering

We have two sites (FNAL and Trieste) that wish to port the monitoring and analysis tools to their site. We are working with the European DataGrid folks to coordinate our efforts.

4 Single Collaborator Efforts and End to End Applications

4.1 ATLAS

4.1.1 US ATLAS Grid Testbed

VDT 1.0 was deployed at testbed sites and the Tier 1 during the period. Testing is underway. Deployment of Magda at a fourth testbed site, UT Arlington, was completed. Magda is now operational at BNL, ANL, LBNL, Boston U, and UTA.

[11] http://www.slac.stanford.edu/comp/net/bandwidth-tests/html/correlation_ACT_PAS.html

4.1.2 EDG Testbed Interface

Through the work of Jerry Gieraltowski (ANL), ATLAS expanded its attention this quarter to establishing an effective ATLAS/PPDG conduit to the EDG working groups to provide feedback on the perceived functionality of the EDG Testbed. Jerry acted as a "new user" of the Testbed to investigate how easy it is to use the functionalities provided by the Testbed. Simple tests were run directly on the Testbed using CERN UI server logon with an INFN certificate. Attempts to run the same set of tests with a valid DOE Science grid certificate have failed so far. Attempts at running a set of simple Globus run tests between a server at ANL and the CE server in the CERN Testbed using the INFN certificate have not been successful so far. Jerry is working closely with Ed May at ANL to construct an EDG UI server following EDG installation instructions. Once complete (expected 4/12), it will be used to test interactions with the CERN Testbed and grid nodes in the US-ATLAS grid. This will involve interactions between the EDG software release and the GriPhyN VDT software release.

4.1.3 Monitoring

Work continued on developing Linux farm monitoring through MDS. A prototype based on the existing local farm monitoring system at the US ATLAS Tier 1 was developed. It can monitor up to 600 nodes, and scalability will be improved in the future. Currently the system can answer limited questions which grid users might ask via MDS. Work is underway with Les Cottrell to import his IEPM monitoring data into MDS. The effort includes creating a MySQL database for the network monitoring data, translating the database schema to the OpenLDAP schema, and implementing the MDS information provider which maps the rows in the relational database to the OpenLDAP object classes. Dantong Yu (BNL) carried out this activity.

4.1.4 Distributed job management

Distributed job management activity, carried on by Pavel Nevski (BNL), focused on completing the ATLAS Data Challenge-directed deployment of a job management infrastructure that integrates the use of distributed data management (Magda) and application metadata management tools. During the period the system operated in production for ATLAS Data Challenge 0 and is now being adapted and extended for use in Data Challenge 1 in the next quarter. Extensions include expanded use of Magda for distributed replication of simulation data, and integration of Virtual Data Catalog work (for automated creation of job specification files) being done by Sasha Vaniachine at ANL. In the next quarter we will begin growing this area as the principal ATLAS PPDG Year 2 focus area, as planned. We plan to begin by examining how we can incorporate usage of MOP.

4.1.5 Data signature

The preliminary design work on event data history performed during the last quarter fed into a major (off-project) effort this quarter in refining the design of an experiment-independent HEP event data store based on a ‘hybrid’ approach in which event data objects are streamed to files while associated metadata is managed in a relational database. Event data history information was incorporated into the design of the hybrid store so as to support data signatures providing comprehensive descriptions and processing histories of all objects in the store. The design will accommodate the use of grid middleware components such as virtual data catalogs to manage job and data descriptions and histories. David Adams (BNL) led and carried out the bulk of this work with participation from V.Perevoztchikov, V.Fine, P.Nevski, T.Wenaus and others. We are now working to inject these efforts into the common project on persistency that is taking shape within the LHC Computing Grid Project.

4.2 BaBar

In addition to the BaBar Database Replication work described above there is effort in cooperation with EU DataGrid. Andy Hanushevsky is collaborating with WP2 in the EU DataGrid on the Giggle Framework to be used for LHC and BaBar to federate various replica catalogs. He attended the 4th DataGrid Conference, March 4th through the 8th. During this time a new concept was developed for managing the Giggle Name Space, called "Bilateral Pass Filters". A paper will be written describing the concept in the next quarter.

In addition, Steve Gowdy and Joseph Perl are participating in the newly formed PPDG activity on Interactive Data Analysis Tools (CS-11).

4.3 CMS

The CMS end-to-end application activities on developing, integrating, deploying and testing a distributed environment for CMS physics event simulation and reconstruction has been described in section 2.3. This effort has been using testbeds and prototype Tier-1 and Tier-2 facilities at Fermilab, UCSD, Caltech, U.Florida and U.Wisconsin at Madison.

Also the work on file replication in collaboration with the European DataGrid project on GDMP has been described elsewhere. At Caltech a new project was started to investigate issues on consistency of replication of states.

Ensuring consistent copying of data is something that cannot be ignored. Data may be copied or re-generated erroneously due to processor or network failing during the replication process, and different copies of data (especially meta-data) may become inconsistent due to receiving different sets of updates or receiving updates in different order. The proposed replication server maintains consistent replicated state in face of processor and network failures. The concept itself was developed through earlier work on the RES. A first prototype of a replicated replica catalogue is close to being finished. As CMS gains experience in distributing its data over Tier-1 and Tier-2 sites, it is expected there will be other areas where consistent state services are required.

Work on the Clarens remote analysis environment continued. The server side of Clarens was “Grid-enabled” to allow authorization and authentication of clients using Grid certificates. It was also adapted to do optional SSL encryption of sessions. Per-session persistent data storage was added using an embedded database. Command line and GUI clients were adapted to make use of these changes and a ROOT-based command-line client development started. Concerning deployment of the Clarens data server, it is currently installed and being tested on the Caltech Tier-2 prototype, as well as on the US CMS testbed machine at Caltech.

The software providing support for certificates for virtual organization (VO), based on a LDAP database, which had been developed at INFN in the context of the European DataGrid project was adapted for use with the DOE science grid certificate authority. The software was installed and tested on the Caltech US CMS testbed and will soon be deployed and used on the whole US CMS testbed.

At Caltech work was performed to understand the requirements for replica catalog and on fault tolerance for grid-wide services. A file-based workload model was developed reflecting an Analysis workload on an LHC experiment grid system.

To investigate possible approaches to serving data for physics analysis work was started on code that allows direct access to CMS physics data stored in an Objectivity database without having to use actual CMS specific libraries. With this code, an CMS database with reconstructed events can be viewed as an Analysis Object Database. Work on writing a web service based on this code is currently underway. Such a web service would allow access to physics event data from simple clients like web browsers. These clients will have access to event data at various levels of detail.

This activity becomes now part of a larger work referred to as the “Caltech LHC Analysis Web Services” (CLAWS), a set of projects aimed to investigate distributed computing and analysis using new technologies. More information on CLAWS can be found in a recent presentation done at CERN CMS Week at this website:

http://documents.cern.ch/cgi-bin/setlink?base=agenda&categ=a02368&id=a02368s5t2/transparencies

The performance of “event serving” for distributed data analysis was investigated using a sample of 180,000 "Tag" events, originally generated from CMS reconstructed data. These demonstrations include the "Bandwidth Greedy" distributed object collection analysis demonstrated at various places, most notably SC2001, and the COJAC Java3D CMS event and geometry viewer. The COJAC tool has two main components: an event reader/renderer and a geometry reader/renderer. The event reader fetches Tag events from an Objectivity database, and displays the particle tracks and jet cones. This code was modified so that

network stress tests could be easily performed. The general idea is to fetch large numbers of Tag objects from a remote database in quick succession. This network stress test as part of COJAC is now in active use on the network connection between Brazil and the US research network. For this purpose, COJAC was installed at UERJ in Rio de Janeiro on a well-connected Linux machine.

Initial prototypes of Web Services for physics analysis were investigated using the Microsoft .NET architecture and tools. The COJAC application was modified so that instead of fetching the Tag objects from a remote Objectivity database, it is now able to fetch Tag objects as binary-encoded SOAP messages from a .NET server running at CACR. This server, running .NET Enterprise Architect, was installed and configured with the latest versions of .NET server and SQLServer.

A simple database table was populated with the Tag event data. Several simple Web Services were implemented in C# to return the requested data as an XML array of integers or as a serialized stream of bytes, which can then be de-serialized easily in clients like COJAC. The Query service allows an arbitrary SQL query to be run against the Tag database. These services can be examined, and used, at: http://baldrick.cacr.caltech.edu/TagServer/TagServices.asmx

Work also continued on the development and tests for the monitoring framework. This was presented at the PPDG meeting in February (http://www.ppdg.net/mtgs/20feb02-toronto/il-ppdg_mon02_02.ppt). The GUI to control the configuration for large systems was improved. This was connected with a SOAP - web service able to export the configuration to clients or other services. GigaSpaces technologies were tested to provide a flexible mechanism for persistent data storage and selection of values based on attributes for the objects. Test of the SNMP base system on a very large farm (lxbatch at CERN with about 400 nodes) was successfully performed.

A network services framework to evaluate the best replication path for a dynamic set of data reflectors on the WAN was developed. The general idea is that each service registers with a set of lookup discovery services and subscribes to remote events generated by other services to notify when a configuration changes or measured parameters vary. It creates a monitoring thread for each possible partner and continuously monitors the current bandwidth to send UDP packets. A simple scheme based on the time to send short and long packets is used to evaluate the available bandwidth. These services are using a lease mechanism to interact with the lookup services and for the events subscription. In case a service does not run properly or has connectivity problems its lease expires and is removed from all the directories. This also generates an event and all the other partners are informed. If a significant change occurs (in configuration or in the measured values) it generates an event and informs all the other partners.

In this way each unit has an up to date picture of the entire connectivity scheme and the effective bandwidth between different nodes. Every service computes a global maximum flow to all the other nodes using a Dijkstra algorithm adapted for this type of problem. The optimization is done each time a new event is received in order to adapt itself dynamically.

Another service was developed which offers a simple GUI to present the entire connectivity for all the services, the evaluated bandwidth between them and the best path for each node. The GUI service discovers all the running units, using the same lookup discovery units, and registers itself for all the remote events signaling a change. When the best graph path is requested for plotting, it contacts directly the unit involved using a proxy published by the services on its web server and gets the Dijkstra minimum path graph.

Tests of these new services were successfully performed between CERN and Caltech.

4.4 D0

We have compiled many use cases to better understand the problem and needs for job submission. A technical overview of existing technology was compiled in an effort to gain a better understanding of the existing grid software and architectures. We are still coordinating manpower outside of PPDG, but have solid contributions from Imperial College, and promising indications from U. Texas Arlington. Rod Walker at IC has worked to understand use of certificates, and contributed a great deal to the test bed and has a CAS configured and running at IC. UTA has been working on the definition of a job, and have produced many use cases to understand the types of processing DAGs anticipated at Dzero.

In the monitoring project Sinisa worked on monitoring use cases for D0 Grid. He spent some time evaluating Globus MDS as possible part of the D0 Grid Monitoring and Information Service infrastructure, and contributed his findings to the D0 Grid Technology Review document. He also worked on the architecture proposal for the D0 Grid Monitoring and Information service. Participated in the PPDG Toronto workshop and presented a talk about status of the D0 Grid monitoring efforts, design requirements and possible architecture for the D0 Grid MIS.

Gabriele Garzoglio introduced GridFTP as a transport protocol for SAM and tested the DOE CA for data transfers with Europe. Worked with Rod Walker (Imperial College) and Alex Flinch (Lancater) for the tests. He is now setting up the test bed to experiment with job submission using the globus toolkit and CondorG. The understanding of the capabilities of these tools is fundamental to take the appropriate decisions when designing and setting up the architecture for remote job dispatch and SAM. The test bed has now three machines at Fermilab; one at Imperial College and, in the near future, UTA

should join with a couple of machines. We've tested job submission from CondorG/GRAM to PBS and Condor; UTA may be able to provide an LSF test system in the near future. During the tests we've exposed a backward incompatibility between the DOE CA certificates and CondorG 6.3.1 and older. Those versions are linked with the old GSI libraries from the globus toolkit 1.1.4. CondorG 6.3.2, which is linked with GSI 2.0, is the version to use with the DOE certificates (http://envisage.es.net/Docs/condorg.txt). He worked with Igor Terekhov, Rod Walker (IC), Jae Yu (UTA) and Tomasz Wlodek (UTA), Doug Olson (LBL) and Michael Helm (LBL), Jaime Frey (Condor) in this endeavor.

Igor and Gabriele are working with the Condor Team, in order to use the Negotiator (which is part of Condor) as the Match Making Service for the DZero Brokering Service. We have agreed on a set of needed features, which should be available in CondorG shortly. A draft of the complete DZero Job and Data Handling architecture can be found at http://www-d0.fnal.gov/computing/grid/ linked from "Architecture (information flow)".

Gabriele worked with John Weigand (Fermilab) and Lee Lueking on the analysis of the load to the Oracle server due to the sam activities of February. This work is published as the PPDG document #14 and was provided to Koen Holtman to on which to base some LHC database transaction rate projections. The document can be obtained from the link "Database Server Oracle Activity Statistics" at http://d0db.fnal.gov/sam/plots-and-stats.html.

Gabriele and Dane Skow (Fermilab) worked on a draft about SAM/PPDG and Security. We agreed on the basic scheme for authentication/authorization and now we are preparing the draft. The same scheme received also input from Igor Terekhov and Rod Walker. We have recently started implementing a prototype of such architecture.

Lee Lueking is working with the networking group at Fermilab to get resources in place to begin running the suite developed and used by IEPM to monitor network performance for Dzero collaborating sites. At Fermilab, Frank Nagy has put up a Linux machine to operate in the DMZ with complete access to the outside network. This is a dual 1 GHz node with Gbit Ethernet, and getting help from Les Cottrell and Connie Logg (SLAC) to deploy the software is underway. Mark Sosebee at UTA will be the first volunteer to get their site up in Arlington for monitoring, and will act as Dzero liaison to work with additional sites. Nearly 2 dozen sites have expressed interest so far, and more will certainly join as the effort develops. We have already begun an initiative to deploy and systematically test large scale data movement to and from these remote sites by having sam station installations at each. This has been quite successful, although our monitoring tools are still being developed to understand the significance and patterns of this traffic.

A serial interface for Root to work with SAM was developed. Gabriele worked with the Root team, in particular Philippe Canel to do this. It has been tested and integrated into the Dzero code base and is being used by some physicists for their root data analysis.

4.5 JLab

Jefferson Lab has continued to test the use of web services in developing and deploying production data grid services. All previous work using raw XML services has now been converted to industry standard SOAP web services, and sufficient services have been completed to begin production testing.

Three web services, one background client, and one user interface client are now ready for testing: 1) a Replica Catalog, 2) a Storage Resource Manager, 3) a File Transfer Service, 4) a File Transfer Spider, and 5) a Grid File Manager user interface. The interfaces for the 3 web services are described using the standard Web Services Description Language (wsdl) at http://lqcd.jlab.org/wsdl/. These interfaces represent a first version of these web services, and will be iterated with others within PPDG and the wider grid community, hopefully resulting in community standard interfaces (see section 2.5).

In addition, the existing disk and silo management software used at Jefferson Lab, JASMine, is in the process of being converted to a SOAP server. Jefferson Lab is thus working with two versions of SOAP SRM's: one adhering closely to the PPDG SRM 1.0 specification (JASMine), and one incorporating extended features (servlet based). The servlet based SRM primarily supports grid access to user managed disk space, and does not include cache and silo management capabilities, and will provide those services by calling the JASMine SRM (thus acting as a proxy).

The servlet SRM supports the features defined in the PPDG document, but includes additional features needed or desired to support the two client programs (spider and GUI). The most notable extensions include a (multi-level) list capability to support a graphical grid file manager, and translation between SURLs (Site URL) and local path names. A configuration option allows a user's home directory to be exported to the grid (accessible only to him), with standard X.509 authentication and mapping to restrict access.

The servlet SRM supports multi-protocol TURL (Transfer URL) negotiation. Jefferson Lab has already integrated (as a demonstration exercise) http, ftp, https, and jparss (java parallel secure stream), and will soon undertake to integrate gridftp.

A separate File Transfer Service (which may eventually be absorbed into the SRM service) supports persistent, queued 3rd party file transfers. It negotiates with the remote SRM for a transfer protocol, and forks the appropriate file transfer client. Protocols can be marked as read-only (pull) in the configuration file.

The Grid File Manager version 0.2 provides a rich and user-friendly web interface for users to manage their files on the data grid nodes (SRMs). This application uses the grid web services via SOAP over https. It is a pure java application, and can be deployed and launched from a web browser by SUN's Java Web Start tool. This allows us to support desktop clients without having to pre-install any grid software, and software updates are automatic. For file transfers between a grid node and the user's desktop it currently supports jparss, and will soon be upgraded to multi-protocol capabilities.

During this period we have also designed and built a Replica Catalog web service that is consistent with the functionality discussed within PPDG. This application features an SQL database backend that collects and stores information about data available across the grid and has been tested with datasets as large as 1 million files.

As a client to the above services, we have implemented a Transfer Spider that will monitor a specified directory and initiate file transfers and replica registration when new files are created. The first version of the spider runs as a special user which allows it to make entries in the Replica Catalog on behalf of other users, so that no user proxies are needed (like an administrator within the Replica Catalog). This allows us to support non-grid enabled batch jobs which simply deposit their output files into the appropriate, watched, directory.

In addition to these two java client applications, we have also implemented browser access to the replica catalog and data grid node web services through translating servlets. These servlets can be demonstrated by starting from http://www.jlab.org/hpc/datagrid/.

These products are being tested at Jefferson Lab and will soon be deployed to MIT for extended testing by a small number of users. The first test will be a production run of a few hundred batch jobs at MIT, with resulting data files automatically registered in the Replica Catalog and replicated to Jefferson Lab.

4.6 STAR

A major event for STAR/PPDG this quarter was the hire of Gabriele Carcassi, funded 70% by PPDG. Gabriele started at BNL on March 25 and has since then mainly investigated Globus and MySQL integration with the aim of steering the STAR database replication using GRID ftp and authentication. Eventually this will be integrated with the STAR file/replica catalog. The rough action plan is:

1. Create a (java based) DB access GUI which provides and centralizes DB management, for a start mainly as required by STAR DBs. This will also allow for consistency checks with the replica.
2. Trigger replication using (1.)
3. Provide an interface to the analysis code, e.g. for the decision which mirror to use.
4. Fully integrate MySQL and Globus authentication.

Dantong Yu and Jerome Lauret have planned and purchased a new star grid-node with Gigabit Ethernet and an overall more powerful server architecture:

- 2 x PIII 1.4 GHz
- 1 GB memory, experimenting with MXT
- 2 x 70 GB SCSI local storage, to be used as HRM buffer (avoid NFS dependence)
- Gigabit Ethernet

The node is currently being configured by Dantong. Pending some issues of incompleteness of Globus 2.0 (official release, missing replica catalog) the new node is expected to be online well in time for us to profit from the BNL/ESNET upgrade to OC12. This is expected to be completed within, at most, four months from now.

Further tests where performed of the BNL/PDSF transfer using HRM. A hardware (now Cisco PIX 535) and software (OS v6.x) upgrade of the BNL perimeter firewall removed the problem of limited (to 64 kByte) buffer-sizes. The same upgrade will soon be finished for the BNL/RCF firewall, pending an RCF-NFS problem. Preliminary tests show that we can now reach a peak transfer rate of 90 Mbit/s and, since we are sharing the bandwidth with entire BNL, are thus limited by the current OC3 connection. These tests where done using Globus 2.0beta, including GRID-FTP.

The stargrid01 node at BNL and the grid (HRM) machine at PDSF both now accept the DOE and EDG (we have French collaborators) certificates and also hold a DOE host-certificate.

Eric Hjorts main activity during this quarter has been learning to use SRM's to transfer data from BNL to LBL. The SRM's themselves work well, however the achievable bandwidth is only a few MB/s (compared to 8-10 MB/s using bbftp). This is a result of two problems: first, a bug in gridFTP, and second, problems with pftp at RCF. The gridFTP bug has been fixed by the Globus people and we are awaiting an official release. The pftp problem at RCF has not been fixed but people there are aware of the problem. An interface to the STAR file catalog is being developed. Other PPDG-related activities include getting a DOE Science Grid certificate and making it work for data transfer, teaching a user to do file transfers with HRMs and developing scripts to get data from disk at RCF instead of from HPSS.

4.7 ANL – Globus

4.7.1 Coordination

As part of an on-going effort for more effective coordination within the Globus project with our collaborators, Jennifer Schopf has taken over the roll of coordinating with in PPDG. She will continue to be assisted in this by Ian Foster, Bill Allcock and Mike Wilde, and by other technology leads of the Globus Project.

4.7.2 Assessment

Members of the Globus team, in order to assess how best their efforts could aid the PPDG experiments, defined the following table of interactions. The following matrix shows two things: which Grid technologies we believe the current experiments are using (a $ for using the Globus solution for the technology, a @ for a different solution), and how applicable we think each technology is to the experiments' requirements (a 1 indicating definitely applicable, 2 – probably applicable, 3 – less applicable or still under development)

	CMS		ATLAS		JLAB*		STAR*		BaBaR*		D0*	
GSI	**$**	1	**$**	1		1	**$**	1	**$**	1	**$** (only w/gridftp)	1
GridFTP	**$**	2	**$**	1		1	**$**	1	**$**	1	**$**	2
Reliable File Transfer		2+		2+		2+		2+		2		
Comm. Auth Service		3 (gridftp)		3 (gridftp)								CAS API
GRAM (job submission)	**$**	1	**$**	1	**@**	1	**@**	1	**@**	1	**@**	1
Rep Cat/ Rep Manager	**$**	3	**$**	1	**@**							
Rep Loc. Svc		3		3	**@**	2		2		2		
MDS	**$**	3	**$**	2		2					**?**	3
Monitoring		3+	**$**	3+		3+		3+	**@**	2		3

* - Experiment is in production today

Using this assessment of current status we identified several deployment goals.

1. Use of GSI and GridFTP across all experiments.
   There had been a successful adoption of GSI and GridFTP across all of the experiments (except JLAB), thereby meeting one of our primary goals. For these two technology focuses, ongoing support of the adoption of the ESNet CA/RA has continued under the guidance of Von Welch. In addition, plans for a full evaluation of the GridFTP performance are planned for Summer 2002.
2. Use of full toolkit by ATLAS and CMS.
   We continue to participate in the planning and deployment of the ATLAS and CMS testbeds participating in both telecons and metings. GRAM is currently in use by both experiments. MDS (and the associated monitoring infrastructure) are being used by ATLAS, and its use in the CMS testbed is planned, with an on-going evaluation of the need for information providers for specific sensor data. Both experiments are using the replica catalog infrastructure through GDMP.
3. Support for Globus Toolkit software all experiments
   We continue to provide support for the use of the Globus Toolkit to all of the experiments associated with PPDG.

4.7.3 Planned alpha-deployments with experiments

We have also identified the following, more focused development and deployment plans

1. RFT
   Work continued on the development of the prototype Reliable File Transfer Service. This service is currently being tested within Globus. We plan to have an alpha-version based on the OGSA

infrastructure available to the PPDG community in July 2002. Detailed information is available at http://www-unix.mcs.anl.gov/~madduri/RFT.html

2. CAS for CMS
   There was an Alpha release of the Community Authorization Service (CAS) enabled GridFTP server. CAS allows resource providers to specify course-grained access control policies in terms of communities as a whole, delegating fine-grained access control policy management to the community itself. Resource providers maintain ultimate authority over their resources. Additional information on CAS can be found at http://www.globus.org/Security/CAS/.
   Initial discussions with CMS (Rick Cavanaugh) have taken place, and they are currently working on an alpha-deployment with Von Welch.
3. RLS
   Work continued on the prototype of the Replica Location Service (RLS), a distributed service for maintaining mappings between logical file names and physical file locations. Functional testing by both the Globus group and EDG WP2 took place, and we believe this will be ready for an outside experiment to alpha test in late April, 2002. Additional information is available at http://www.isi.edu/~annc/RLS.html.
4. MDS Information providers
   We are currently in discussion with members of the CMS testbed and from BaBar in order to extend the range of information providers supplied with the Globus Monitoring and Discovery Service (MDS). This is related to our on-going efforts with the joint PPDG/GriPhyN/iVDGL monitoring project, detailed in that section.

4.7.4 Talks and Publications

We conducted two in-depth Developers and Administrators tutorials. The tutorials were a tremendous success with over 150 attendees in Edinburgh, Scotland and over 200 at Argonne National Laboratory in Chicago, IL. The tutorials covered all aspects of software development with the Globus Toolkit including presentation, code walk-throughs, and demonstrations. The Argonne tutorials were recorded and are being indexed for streaming playback over the web.

4.8 NERSC – SDM

People involved: Junmin Gu, Alex Sim, Alex Romosan, Arie Shoshani

There were 3 activities during the last quarter that we can report progress on:

1) Installation and testing of HRM at BNL and NERSC/PDSF

This activity involved installation of HRM and associated software in two locations: BNL and NERSC/PDSF. This installation required extensive coordination with system people at both locations. The installation included DRM, TRM, and HRM-CLI (the client interface program to HRM), as well as the latest version of the Globus Toolkit gt2-beta1.

We used the HRM installation to perform several tests. By plotting the log information generated by HRM, we managed to discover a bug in the Globus-URL-Copy module. This was further followed up with one of the Globus developers, and the problem was fixed. This fix is now included in the newly released Globus Toolkit gt2-beta2. These tests and a representative plot can be seen in slides 29-31 of the presentation:

http://sdm.lbl.gov/srm/documents/02.02.srm.joint.design/index.htm.

2) Stress testing and enhancements to the DRM and the HRM

We conducted extensive stress testing during this period for both DRM and TRM. Several bugs were discovered and fixed. In addition, several previously known bugs were fixed, including core dump with heavy memory usage, deadlocks, and memory leaks. We believe that these products are quite robust at this time. We will verify this in the future with continuous use of the DRM and HRM.

Several new features were added to DRM during this period as well. Event logs were added, and transfer rates are now measured. These are used to provide time_to_service estimates. This feature is used currently by the GridFTP-HRM capability described in the next task. Another feature added is "flexible file size". This feature permits the requested for either getting or putting a file into DRM without the file size specified. This requires the DRM to allocate space dynamically, a feature that is quite complex to accommodate, since the size of the transfer has to be dynamically monitored. Finally, a "state_recovery" capability was added. This capability makes sure that the state of DRM is captured on disk, so that in case that DRM crashes or is stopped, it will recover to the same state when restarted.

3) Developing a new GridFTP-HRM capability

This is a new task that was initiated this quarter. Its goal is to allow regular GridFTP to get/put files into HPSS through the use of HRM. Since HRM is designed to interface to HPSS, having gridFTP use it means that GridFTP does not have to be implemented on HPSS. The idea is to modify the GridFTP server daemon, so that when requests arrive it invokes HRM. HRM then communicates with HPSS to perform the request (either get or put), and then let the FTP daemon continue to process the file transfer. In the case of a "get", the HRM allocates space on its disk, requests the file to be transferred into its disk from HPSS, and then provides the disk address to the FTP daemon so it can transfer to file to the requester. In case of a "put" the HRM also allocates space to its disk, then lets the FTP daemon write the file to the allocated space, and schedules the file to be transferred to HPSS.

During this quarter, we developed the infrastructure to allow the standard GridFTP tools (globus-url-copy) to interface to the LBNL HRM (DRM+TRM) in a transparent manner. This interface (called hrmFTP) consists of a multi-threaded daemon, which talks to a patched version of the GridFTP server (based on globus_gsi-wuftpd-0.6) via shared memory, and translates the ftp commands (STOR (store), RETR (retrieve), SIZE (size)) to appropriate requests to HRM using CORBA. Each thread of the hrmFTP daemon talks to one instance of the ftpd daemon (with each instance of the ftp daemon corresponding to one remote connection). We have tested the hrmFTP using the HPSS facility at NERSC in a simulated multi-user environment (multiple simultaneous requests for storing and retrieving files). During the next quarter, this capability will undergo extensive testing, and is expected to be released.

In addition to the above task the following presentation were prepared and delivered. Arie Shoshani participated in the Grid Forum meeting, and presented at talk at the Data Management Working Group session on "Storage Resource Management (SRM) and GridFTP: how do they interact?" This talk described the considerations and design for the third task described above: developing a new GridFTP-HRM capability. This presentation can be found at: http://sdm.lbl.gov/srm/documents/02.02.srm.joint.design/index.htm. Arie also participated in the PPDG meeting that followed the Grid Forum meeting, and gave a talk on the joint design of SRM representing the results of a coordinated effort to standardize the SRM interface in PPDG as well as in the EDG. This presentation is available at: http://sdm.lbl.gov/srm/documents/02.02.srm.joint.design/index.htm.

4.9 SDSC – SRB

The activities at the San Diego Supercomputer Center in support of the PPDG have focused on

- BaBar support
- Derived data product support
- web services interface
- demos of web services
- description of data grids
- interactions with Chip Watson on service definition

4.9.1 Support for derived data products

It is possible to create collections of the output files that are created by applications that analyze experimental data. The derived data products can be characterized and stored with metadata that allows their discovery. We used the Virtual Data Language developed by the Globus team to describe the creation of derived data products. We built a parser for VDL files, stored the results in the MCAT catalog, and then supported general queries against the VDL parameters and file names.

We developed a prototype in Perl to extract metadata from VDL files and then use SRB data mining tools to save the VDL files as datasets into any remote storage place running SRB and to save the extracted metadata into the MCAT catalog. Then we used mySRB (a SRB web interface) and other command-line tools to search the VDL files based on job information such as application name , input parameters and output file names.

A corresponding web service to mine a VDL file and its metadata was developed based on this Perl program. A comparison of the SRB VDL metadata mining using MCAT with the similar technology developed by the Globus team is under investigation.

4.9.2 Description of data grids

A comparison of seven data grids has been made to understand the key features needed to support access to remote data. The comparison included the Storage Resource Broker (SRB) data grid from the San Diego Supercomputer Center, the European DataGrid replication environment (based upon GDMP, a project in common between the European DataGrid and the Particle Physics Data Grid, and augmented with an additional product of the European DataGrid for storing and retrieving meta-data in relational databases called Spitfire and other components), the Scientific Data Management (SDM) data grid from Pacific Northwest National Laboratory, the Globus toolkit, the Sequential Access using Metadata (SAM) data grid from Fermi National Accelerator Laboratory, the Magda data management system from Brookhaven National Laboratory, and the JASMine data grid from Jefferson National Laboratory.

Key features that were used by the majority of the data grids included: use of a logical name space to build global identifiers, hierarchical organization of the logical name space, addition of attributes to the logical name space for grid parameters, federated client server architecture, use of GSI authentication, etc. The report has been used by each of the participating sites to gain an understanding of the common features, and is promoting the development of additional common features.

4.9.3 Web service definition

The grid comparison was extended to derive a set of common capabilities or operations that were supported by the seven data grids. In turn, these common operations were compared to the SRM interface that is being implemented by Chip Watson. While both approaches are resulting in web service interface definitions, there remain some differences in the approaches. What must be resolved next are access to community versus user-owned data, asynchronous staged access versus synchronous access, server initiated versus client initiated data transfers, specification of the name space on which the operations are done (either global name space, replica catalog name space, SRM logical name space, physical file names), and set of permissible operations on files in the name space.

Demonstrations of the SRB WSDL interface were given at the Global Grid Forum.

4.10 . Wisconsin – Condor

During this time, Wisconsin assumed primary development responsibility for the MOP software system, after the departure of James Amundson at Fermilab.

Wisconsin coordinated with the GriPhyN Virtual Data Toolkit (VDT) developers and used a prototype VDT version as the base software deployment on the US-CMS Testbed.

Wisconsin also collaborated with the EDG WP1 project team to discuss PPDG and EDG architectures, coordinate development activities, and try to ensure future compatibility.

Wisconsin continued to explore distributed error-propagation and reporting issues, presented a talk at January's PPDG focus group meeting at JLab, and prepared a draft paper on the topic.

Wisconsin continued its participation on the HENP InterGrid Joint Technical Board.

There was also work on the Hawkeye system[12] which is being used to monitor the US/CMS testbed.

5 Appendix

5.1 List of participants

TEAM	Name	F	Current Role CS	1	2	3	4	5	6	7	8	9	10
Globus/ANL	Ian Foster	Y	Globus Team Lead, GriPhyN PI, iVDGL, GriPhyN						x	x			
	Mike Wilde	Y	GriPhyN coordinator					x	x				
	Jenny Schopf	Y	GriPhyN collaborator			x							
	Willliam Alcock	Y							x				
ATLAS	Torre Wenaus	N	ATLAS Team Lead.		x			x					
	L. Price	N	Liaison to HICB, HICB Chair										
	D. Malon	N											
	A. Vaniachine	Y											
	E. May	N						x					x
	Rich Baker	N											
	Alex Undrus	Y											
	Dave Adams	Y											
	Wengshen Deng												
	G. Gieraltowski	Y									x		x
	Dantong Yu	Y	Monitoring			x							
STAR	M. Messer	N	STAR Team Lead										
	Eric Hjort	Y					x	x					
CMS	Lothar Bauerdick	N	CMS Team Lead. GriPhyN collaborator										
	Harvey Newman	N	PPDG PI. GriPhyN collaborator, Co-PI iVDGL										
	Julian Bunn	N	CMS Tier 2 manager, GriPhyN & iVDGL collaborator										
	Tokako Hickey	Y	CS-8:Robust Job Scheduling, GriPhyN collaborator								x		
	Conrad Steenberg	Y	CS-8:Analysis Tools, GriPhyN collaborator								x		
	Koen Holtman	N	GriPhyN collaborator										

[12] http://www.cs.wisc.edu/condor/hawkeye/

	Iosif Legrand	N	CS-8:Monitoring Tools								x		
	Vladimir Litvin	N	GriPhyN collaborator	x	x								
	Jim Amundson	Y			x								
	Shazhad Muzzafar	Y						x					
	James Branson	N	CMS Tier 2 manager										
	Ian Fisk	N	CMS Level 2 CAS manager, iVDGL liaison										
Coordination	Ruth Pordes	Y	PPDG coordinator							x			
	Doug Olson	Y	PPDG coordinator				x			x		x	
D0	Lee Leuking	N	D0 Team Lead		x	x							
	Igor Terekhov	Y		x	X	x							
	Sinisa Veseli	Y				X							
	Gabriele Garzoglio	Y							X				
HRM/LBNL	Arie Shoshani	y	SRM Team Lead. GriPhyN collaborator				X						
	Alex Sim	Y					X						
	JunminGu	Y					X						
	Alex Romosan	Y					x						
SRB/UCSD	Reagan Moore	Y	SRB Team Lead. GriPhyN collaborator					X		X	x		
	Bing Zhu	Y	CS-8: Web Services					x			X		
SLAC	Richard Mount	N	PPDG PI, BaBar Team Lead										
	Robert Cowles	N										X	
	Andrew Hanushevsky	Y						x	X				
	Adil Hassan	Y						x	X				
	Les Cottrell	N	IEPM Liaison			X							
JLAB	William Watson	Y	JLAB Team Lead				x	x	x		x		
	Ian Bird	N					X						
	Andy Kowalski	N					X						
	Bryan Hess	Y	CS-8:Web Services				x				X		
	Ying Chen	Y	CS-8: Web Services				X	x			x		
	Walt Akers	Y	CS-8, Web Services					x			x		
Condor/U.Wisconsin	Miron Livny	Y	PPDG PI, PPDG Coordinator. GriPhyN collaborator	x	x	x	x		x		x		
	Paul Barford	Y				X							
	Peter Couvares	Y		X	x								
	Rajesh Rajamani	N			x						X		

5.2 Appendix - PPDG Meetings

Monday, January 7, 2002, (times are US Pacific zone)	
7 a.m. - 8:30 a.m.	**HICB JTB Phone Con**
1 p.m. - 3 p.m.	**Replication Focus teleconference**
Thursday, January 10, 2002, (times are US Pacific zone)	
9 a.m. - 5:30 p.m.	**Robust Replication Focus Meeting, JLAB** **URL:** http://www.ppdg.net/mtgs/10jan-02/agenda.htm
Monday, January 14, 2002, (times are US Pacific zone)	
1 p.m. - 3 p.m.	**Replication Focus teleconference**
Monday, January 21, 2002, (times are US Pacific zone)	
1 p.m. - 3 p.m.	**Replication Focus teleconference**
Wednesday, January 23, 2002, (times are US Pacific zone)	
12:30 p.m. - 2:30 p.m.	**PPDG steering committee** **URL:** http://www.ppdg.net/mtgs/Jan23-02.htm
Thursday, January 24, 2002, (times are US Pacific zone)	
2:30 p.m. - 4 p.m.	**PPDG-GriPhyN-iVDGL Joint Monitoring Working Group** **URL:** http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/
Wednesday, January 30, 2002, (times are US Pacific zone)	
12:30 p.m. - 2:30 p.m.	**PPDG weekly phone meeting** **URL:** http://www.ppdg.net/mtgs/phone/020130/default.htm
Monday, February 4, 2002, (times are US Pacific zone)	
7 a.m. - 8:30 a.m.	**HICB JTB Phone Con**
Wednesday, February 6, 2002, (times are US Pacific zone)	
12:30 p.m. - 2:30 p.m.	**PPDG weekly phone meeting** **URL:** http://www.ppdg.net/mtgs/phone/020206/default.htm
Thursday, February 7, 2002, (times are US Pacific zone)	

12:30 p.m. - 2 p.m.	**PPDG-GriPhyN-iVDGL Joint Monitoring Working Group** **URL:** http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/

Wednesday, February 13, 2002, (times are US Pacific zone)

12:30 p.m. - 2:30 p.m.	**PPDG weekly phone meeting** **URL:** http://www.ppdg.net/mtgs/phone/020213/default.htm

Wednesday, February 20, 2002, (times are US Pacific zone)

12 p.m. - 8 p.m.	**Collaboration meeting, Toronto** **URL:** http://www.ppdg.net/mtgs/20feb02-toronto/agenda.htm

Thursday, February 21, 2002, (times are US Pacific zone)

All Day	**Collaboration meeting, Toronto** **URL:** http://www.ppdg.net/mtgs/20feb02-toronto/agenda.htm

Wednesday, February 27, 2002, (times are US Pacific zone)

12:30 p.m. - 2:30 p.m.	**PPDG steering meeting**

Thursday, February 28, 2002, (times are US Pacific zone)

10 a.m. - 12 p.m.	**PPDG Site-AA discussion**

Monday, March 4, 2002, (times are US Pacific zone)

7 a.m. - 8:30 a.m.	**HICB JTB Phone Con**

Wednesday, March 6, 2002, (times are US Pacific zone)

12:30 p.m. - 2:30 p.m.	**PPDG weekly phone meeting** **URL:** http://www.ppdg.net/mtgs/phone/020306/default.htm

Thursday, March 7, 2002, (times are US Pacific zone)

2:30 p.m. - 4 p.m.	**PPDG-GriPhyN-iVDGL Joint Monitoring Working Group** **URL:** http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/

Wednesday, March 20, 2002, (times are US Pacific zone)

12:30 p.m. - 2:30 p.m.	**PPDG Steering Committee meeting** **URL:** http://www.ppdg.net/mtgs/Mar20-02.htm

Thursday, March 21, 2002, (times are US Pacific zone)

2:30 p.m. - 4 p.m.	**PPDG-GriPhyN-iVDGL Joint Monitoring Working Group** **URL:** http://www-unix.mcs.anl.gov/~schopf/pg-monitoring/

Monday, April 1, 2002, (times are US Pacific zone)

7 a.m. - 8:30 a.m.	**HICB JTB Phone Con**

Wednesday, April 10, 2002, (times are US Pacific zone)

12:30 p.m. - 2:30 p.m.	**PPDG weekly phone meeting** **URL:** http://www.ppdg.net/mtgs/phone/020410/default.htm

Wednesday, April 17, 2002, (times are US Pacific zone)

12:30 p.m. - 2:30 p.m.	**PPDG weekly phone meeting** **URL:** http://www.ppdg.net/mtgs/phone/020417/default.htm

Thursday, April 18, 2002, (times are US Pacific zone)

4 p.m. - 5 p.m.	**PPDG IDAT Wshop pre-meeting call**

Friday, April 26, 2002, (times are US Pacific zone)

All Day	**PPDG SiteAAA meeting, ANL** **URL:** http://www.ppdg.net/mtgs/26apr-02-siteaa/agenda.htm